# #So Berlin?

# Where to move within Germany, if you want to enjoy the night while caring about rental prices and wages?

Clustering Germany's biggest cities according to their nightlife, average income and rental prices

Felix Klapprott

Mai, 2020

## Part II: Data

For my analyses, I will use data on Income, Rental Prices and a city's nightlife.

I decided to focus the analyses on cities with more than 500 000 Inhabitants. A useful overview is found freely available on the German Wikipedia at https://de.wikipedia.org/wiki/Liste_der_Gro%C3%9F-_und_Mittelst%C3%A4dte_in_Deutschland . An extracted list will be the basis of all analyses as it provides a list of official names and the needed selection of all German cities with more than 500.000. The list will be used to find the names of all cities with more than 500.000 Inhabitants in order to find out more about them in later steps (i.e. using official statistics (rental prices / income), geo-data (visualization) and foursquare data (nightlife) )

The foursquare API will provide me with useful information about the type of venues in a given city. I will retrieve information about relevant venues for all of the cities with more than 500.000 inhabitants. I will focus only on venues which are registered as "Nightlife spot" and plan to look for a specific pattern of different venues rather than the mere number of options. I will calculate relative frequencies of different venue types per city to find similarities between cities, i.e. for example distinguishing cities where you would find mainly cocktail and wine bars as opposed to night clubs, pubs and beer gardens.

Additionally rental prices for these cities are needed. Rental prices will be used as a variable in the later cluster analysis to find patterns among Germanys biggest cities (i.e. are their overlooked and remarkably cheap yet interesting cities in Germany?). The data set (German) including a comparison to 2009 is accessible (premium/university access) at https://de.statista.com/statistik/daten/studie/167163/umfrage/mietentwicklung-in-den-deutschen-grossstaedten/ . As it is not publicly available I will download it and convert to csv in order to work with it in Python. The data is based on 234,000 offers advertised on the online platform immowelt.de. The prices reflect the median of the rental apartments and houses offered in the first half 2019. As many students decide to share flats to benefit from old contracts with cheaper prices these data is only

an approximation of true costs of living. Nevertheless it will give an impression of whether a city is rather expensive or cheap to live in.

Average Income will be used as a variable in the later cluster analysis to find patterns among Germanys biggest cities and to distinguish cities based on economic potential for later graduates, as this might also influence their decision. The most recent data I could find on accessible income per capita in German cities was collected in 2016. The data set can be found on https://de.statista.com/statistik/daten/studie/998971/umfrage/verfuegbares-einkommen-in-den-groessten-staedten-in-deutschland/ . As it is not easily available I will download it too and convert to csv in order to work with it in Python.